

2006 XSD Scientific Software Workshop Report

**Kenneth Evans, Jr., Francesco De Carlo,
Pete Jemian, Jonathan Lang, Ulrich Lienert,
John Maclean, Matt Newville,
Brian James Tieman, Brian H. Toby,
Michel van Veenendaal**

October 31, 2006

Contents

Executive Summary	3
Key Findings and Recommendations	4
Detailed Findings and Recommendations	5
Appendix 1: Charge to the Committee	11
Appendix 2: Summary of the User Community Survey	12
Appendix 3: Summary of Workshop Presentations.....	14

Executive Summary

In May of 2006, a committee was formed to assess the fundamental needs and opportunities in scientific software for x-ray data reduction, analysis, modeling, and simulation. This committee held a series of discussions throughout the summer, conducted a poll of the members of the x-ray community, and held a workshop. This report details the findings and recommendations of the committee.

Each experiment performed at the APS requires three crucial ingredients: the powerful x-ray source, an optimized instrument to perform measurements, and computer software to acquire, visualize, and analyze the experimental observations. While the APS has invested significant resources in the accelerator, investment in other areas such as scientific software for data analysis and visualization has lagged behind. This has led to the adoption of a wide variety of software with variable levels of usability. *In order to maximize the scientific output of the APS, it is essential to support the broad development of real-time analysis and data visualization software.* As scientists attack problems of increasing sophistication and deal with larger and more complex data sets, software is playing an ever more important role. Furthermore, our need for excellent and flexible scientific software can only be expected to increase, as the upgrade of the APS facility and the implementation of advanced detectors create a host of new measurement capabilities. New software analysis tools must be developed to take full advantage of these capabilities. *It is critical that the APS take the lead in software development and the implementation of theory to software to ensure the continued success of this facility.*

The topics described in this report are relevant to the APS today and critical for the APS upgrade plan. Implementing these recommendations will have a positive impact on the scientific productivity of the APS today and will be even more critical in the future.

Key Findings and Recommendations

- F1:** *There is a lack of a coordinated scientific software development at the APS.*
- R1:** The APS should make a commitment to establishing a scientific software development group. This group would have a twofold mission. First, it would assist in integrating existing analysis and modeling codes with the data acquisition software. Second it would develop new codes for both existing and new techniques. The committee recommends that 2 FTEs should be tasked with software development and an additional 2 FTEs for data acquisition in FY2007. The new scientific software staff would nucleate an eventual group of 10-14 scientific software developers.
- F2:** *The flow of data from the acquisition system into analysis codes is inefficient for many experiments currently performed on APS beamlines.*
- R2:** Data acquisition, analysis, modeling, and simulation must become better integrated. Staff effort needs to be dedicated to improve communication links between various software packages. The APS should develop a workflow application to manage the data flow from acquisition through modeling and analysis. Collaboration with the neutron community, which is developing similar workbench programs along with standardized data formats, should be explored.
- F3:** *There is a serious need for tools for scientific visualization.*
- R3:** The APS should provide a common tool set for viewing 1-D, 2-D, 3-D, and higher dimensional data. This would improve the ability of the experimenter to evaluate the quality of the data in real time at the beamline and aid the user in viewing and analyzing data at the user's home institution. This software should also be available to APS users for use at their home institutions.
- F4:** *There has been a lack of high-performance computing in places where it is essential to perform data reduction and analysis at the APS.*
- R4:** Computer grids and clusters have been applied to tomography reconstructions, but many other techniques involving the reduction of large 2-D data sets, such as XPCS, micro-diffraction, and stress/stain measurements, could benefit significantly from the use of such advanced computing. We strongly encourage the APS to set up a local high-performance computing cluster and develop software to apply such a resource to techniques beyond tomography.
- F5:** *The limited theoretical support reduces the scientific output and impact of the APS.*
- R5:** The APS should strengthen the theory group by establishing a visiting scientist program to bring in experts in both theory and data analysis to the APS. The longer-term plan should also include increasing the number of permanent theorists in the group. The synergy of these scientists with APS researchers and with the scientific software development group would help facilitate the conversion of code produced elsewhere into user-friendly software on the beamlines.

Detailed Findings and Recommendations

Software is essential to improve the scientific output of the APS. Today a significant fraction of the data taken cannot be evaluated in real time but only months after the data collection. A lack of real-time data-reduction procedures and visualization tools prevents scientists from evaluating data quality while measurements are still in progress, leading to inefficient instrument use. The software currently in use is a patchwork of unrelated programs that do not work together easily and are not optimized to the needs of APS users and staff. Many of the programs have been developed by individual scientists but are not maintained, portable, or open source. Furthermore, there is a lack of tools to allow users to process the entire data work flow from acquisition to publication, encompassing the ability to modify the experiment in real time after seeing the results of modeling, analysis, and visualization. *It is essential for the APS to take a more concerted and focused approach to the development of real-time analysis and data visualization software than has been done to date.*

By taking a leading role in the development and maintenance of open-source software for the entire process of x-ray data analysis, and by partnering with APS users and with scientists and developers at universities, national laboratories, and other scientific facilities throughout the world, the APS has a unique opportunity to improve the state of the art of x-ray science.

Scientific Software Development Group

The first step toward addressing the software development issues for the APS should be to establish a scientific software development group. Such a group is essential and required to provide the necessary coordinated software development. A dedicated group will additionally provide a center of responsibility and lead to a more efficient and planned effort. It would also provide direction and expertise and be a focal point for collaboration with other development here, at universities, at other institutions, and in other scientific communities. The members of this group should be directly involved with beamlines to ensure a close collaboration with the APS scientific staff.

In particular, this group would: (1) Combine existing analysis and visualization codes with beamline data acquisition software and transform these codes into easy-to-use software. (2) Provide a scientific workbench program that is easy to use and learn and from which users can access all the software that is necessary to manage the entire scientific work flow. (3) Create new analysis and visualization applications that can be used on all beamlines and that are easily integrated into the standard workbench. (4) Develop a software framework, perhaps more than one, that provides tested and debugged scientific routines, such as fitting and visualization, which can be used by developers to create applications. (5) Create an interface to the facilities necessary to provide high-performance computing. (6) Provide documentation, distribution, maintenance, and support.

Scientific Workflow

In accessing software needs it is valuable to think of the entire scientific workflow process. This process is shown schematically in Fig. 1. Different parts of this workflow will be described in the following sections.

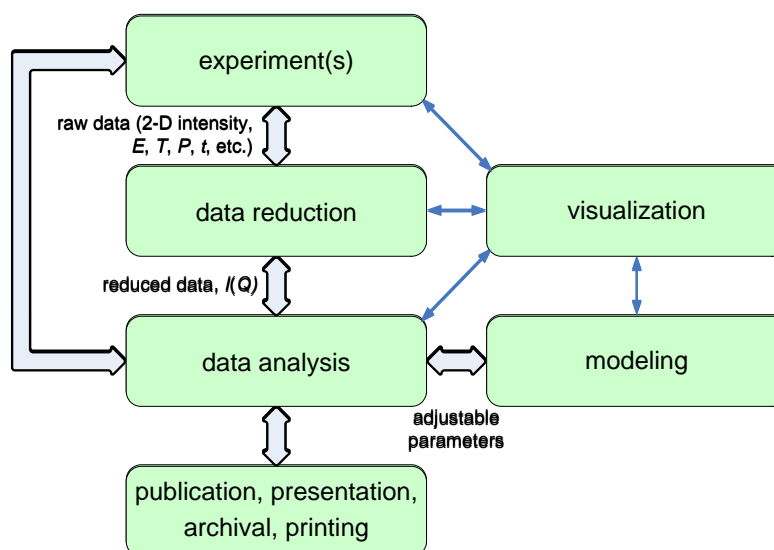


Figure 1. Scientific workflow diagram.

Data Handling

Much of the newest science requires faster data acquisition and storage. The advent of new 2-D detectors has greatly increased the capability to produce more data at ever-increasing rates. Some experiments are now producing as much as 1 Tbyte/day. As brighter sources and faster detectors become available, these data rates will only increase. Software must either be developed by the x-ray facility or obtained from the user communities. Other communities with experience in handling large quantities of data are high-energy physics and astronomy. A means of both storing and rapidly retrieving these data, especially for comparison with modeling and analysis programs, must be available. It is particularly critical to be able to reduce large data sets into viewable scientific results.

In addition to traditional backups, a data archival system needs to be implemented. Currently, many beamlines store archives of data on tapes in cabinets or drawers making it difficult, if not impossible, to retrieve the data at a later date. Data needs to be cataloged and indexed for easy identification and retrieval in order to make it easier to correlate results from differing methods and to share data with collaborators.

Common Data Format

The data generated by the x-ray community would benefit greatly from a widely accepted, common data format. Such a common format would facilitate reuse of analysis and visualization codes, foster interoperability between similar types of instruments within the APS, and enhance workflow design for data analysis. We understand that the move to a common data format will not be easy, and part of the process will be to create and demonstrate to the community the many benefits worthy of the investment. The neutron community has developed such a standardized data format, NeXus, which is gaining acceptance, but NeXus has made little headway into the synchrotron community. Significant amounts of work are needed to develop templates for many common x-ray data collection instrument configurations to integrate the NeXus format into current x-ray data acquisition and analysis tools. While our current users do not yet appreciate the benefits to be gained from data formatted in NeXus, this is expected to change as DANSE provides tools for the SNS and reaches out to the APS.

Support at the APS for a common data format, as well as internal expertise in its implementation, would be beneficial to the community and help maintain APS's leadership role in x-ray research. The APS should encourage the use of NeXus within new software tools. In the interim, efficient, user-friendly conversion routines should be developed for putting data into and extracting it from NeXus files.

Visualization

One of the most pressing issues in scientific software is online visualization and processing of data during collection at the beamline. Visualization aids in every phase of the scientific workflow, as indicated in Fig. 1. Many tools exist for these tasks, but the lack of the implementation of standard tools at the APS has caused our users to identify problems in data collection long after the experiments are over. This both creates a training hardship for users and prevents them from identifying potential problems with the data acquisition in a timely manner. Visualization software at the APS should be developed both in addition to and in support of commercial visualization environments. There is a critical need for the APS to make this visualization software available to users, so that they may also perform offline analyses at their home institutions.

Modeling, Analysis, and Theory

Theory and modeling represent an underutilized resource at the APS. Experimenters would benefit greatly if they could simulate their proposed measurements before performing experiments to determine the key experimental parameters to measure or to test if the intended measurement will discern between competing models. An appreciation of the value of theory and modeling in data interpretation has recently been supported by a BESAC subcommittee report on theory and computation.¹ This report urged that the BES “significantly augment its theoretical and computational programs

coupled to experimental research at its major facilities.” Enhancing such support here at the APS would mesh nicely with these efforts.

The quantitative understanding of x-ray data relies on sophisticated numerical modeling of the data in terms of simple physical and chemical models describing the system. Here, the need for robust, state-of-the-art analysis and fitting programs is very important for the APS users, and indeed x-ray scientists around the world. In addition, a significant amount of the x-ray data collected at the APS is limited by the ability to simulate and predict the measured spectra to the measurement accuracy. It is therefore important to bring together modern analysis algorithms and *ab initio* theoretical work at the APS, including both in-house development and synthesis of work done outside the APS.

Complex materials, such as disordered materials, catalysts, and host-guest systems, present a particular challenge for data fitting in that a single measurement may not present enough information to uniquely determine a model. It is currently possible to fit a model simultaneously to both x-ray and neutron diffraction data, but a growth direction for materials research will be to combine use of many other types of spectroscopic and scattering measurements. This cannot be done at present. Any measurement that can be quantitatively simulated from an atomistic model can be used for model fitting if the appropriate software is created.

The challenge to the APS is to develop the most fruitful interactions with theorists. The presence of a strong theoretical group is essential in interfacing with the theoretical community outside the APS and in providing support to users and developers. Furthermore, mechanisms such as summer and sabbatical support for visiting experts in both theory and data analysis can be used to further leverage our interactions with outside communities. The synergy of these scientists and the recommended scientific software development group would help facilitate the conversion of code developed here and elsewhere into easy-to-use experimental modeling and analysis software at the beamlines.

Real-Time Analysis and Experimental Feedback

Real-time analysis of the experimental data, using modeling and analysis software, is essential in order to make the most efficient use of beam time. Many times, the usefulness of an experiment can only be determined once the data are processed. This is particularly true of experiments that require the reduction of large 2-D data sets, such as tomography, protein crystallography, microdiffraction, stress/strain measurements, x-ray photon correlation spectroscopy, etc. This need is only expected to grow as new types of 2-D detectors are brought online. Software needs to be developed to allow a user to efficiently complete data reduction, visualization, and in many cases a first cut at the analysis and modeling during the data acquisition process. This would allow users to change the experiment appropriately, as shown by the feedback arrow in Fig. 1.

Quick real-time analysis would also alleviate many difficulties in transporting large volumes of data back to a user's home institution. It is imperative that the software tools necessary to quickly process the data exist and be readily available at APS beamlines. In many instances, this means having high-performance computing systems directly coupled to the data acquisition systems and available for real-time data analysis and feedback. Furthermore, the APS should pursue methods to enable users to access these same analysis tools remotely from their home institutions, if further data analysis is required.

High-Performance Computing

The rapid pace of data acquisition at the APS relies on state-of-the-art data reduction, analysis, modeling, and theory. There are many scientific calculations for which high-performance computing facilities are necessary. As one example, making density functional theory calculations readily available to APS users will greatly enhance their ability to simulate atomic structures of complex materials, which is of great value in spectroscopic analysis.

It is also important to do these calculations quickly, especially if the results will be used to feed back into an experiment in progress. Argonne's excellence in high-performance computing is a key asset to the staff and users of the APS. However, the scientific software group must provide expertise in adapting software for high-performance computing to foster usage of these capabilities.

A critical difference between traditional high-performance computing and the computing required for a facility such as the APS is the need for dedicated real-time access to the computing resources. Traditional techniques such as queuing of jobs that are aimed at maximizing utilization of CPU cycles need to be rethought, as experiments need on-demand access to computing power in order to see the results as soon as possible to further guide the experiment.

Workbenches, Frameworks, and Common Applications

Workbenches and frameworks are software concepts in use by many communities such as neutron scattering, high-energy physics, and astronomy. A workbench is an application that helps organize a person's work, allowing him to perform most or all of the tasks involved in his workflow. A framework is a set of routines that can be used to build applications. These are software concepts that are not specific to the x-ray community, but which could be used to advantage in x-ray software.

The APS needs a workbench application to integrate and manage an entire process from acquisition through analysis and modeling, including after-analysis steps, as pictured in Fig. 1. This application should be easy for new users to learn and powerful enough to provide advanced users access to their modeling and analysis software in real time. A common workbench platform for both x-ray and neutron analysis would be

beneficial, as many users collect data at several facilities. In particular, collaboration with the neutron community, which is also developing similar workbench programs, should be explored.

There is a need for software applications to assist users and beamline scientists in those tasks that are common to many experiments. This may be done through new development and coordination of existing efforts. Two of the goals are to provide assistance and to reduce duplication of effort. In addition, a framework of routines that addresses common needs in data analysis would greatly increase the effectiveness of x-ray software development both by the scientific software designers and by beamline staff involved in software creation. The scientific software group should investigate existing open-source frameworks, develop additional frameworks as necessary, and provide the required support and consultation.

References

1. "Opportunities for Discovery: Theory and Computation in Basic Energy Sciences," January 2005,
http://www.science.doe.gov/bes/besac/Theory%20and%20Computation_rpt.pdf

Appendix 1: Charge to the Committee

The committee was charged by Gabrielle Long, Division Director, APS X-ray Science Division, to organize and convene a full-day or half-day workshop in August 2006 at the APS to determine our fundamental needs and opportunities in scientific software systems for x-ray data reduction, analysis, modeling, and simulation.

1. What is the current status of scientific software availability at APS beamlines?
What scientific opportunities are addressed by the current software?
2. How would optimal scientific software and analysis systems improve the scientific impact of APS? Can we prioritize APS data analysis software needs?
What new analysis tools will be needed in the future?
3. What is the best way to satisfy the demand for improved scientific software?
What is the optimum path for software development efforts within XSD to achieve the objectives?

Appendix 2: Summary of the User Community Survey

A survey was emailed to the APS user community in June. The committee received 22 responses from users in a variety of disciplines. All the responses are contained in the 2006 XSD Scientific Software User Survey [A1]. A summary of the most common themes found in the responses is given below.

- **Reduction and real-time analysis of 2-D data**

The most common software need mentioned in the survey responses was the inability to quickly reduce large 2-D data sets into viewable results. This was mentioned by over half the respondents in regard to a diverse set of techniques such as microdiffraction, stress/strain measurements, XPCS, and white-beam Laue diffraction. Routines for quickly and easily transforming diffraction images to reciprocal space images were requested. Clearly the lack of such software tools for dealing with 2-D data is a current limitation on the scientific output of the APS.

- **Data visualization tools**

The need for better data visualization tools (particularly for 2-D and higher dimensional data) was also cited by many of the respondents. Specifically, the ability to customize the data visualization programs for different types of experiments was requested. The need for a simple scriptable programming environment to be able to perform more routine data manipulation was identified as a particular need.

- **Modeling, Analysis, and Theory**

Most modeling, analysis, and theory code is developed by scientists here at the APS and at outside institutions by people who are not software experts. Several of the responders requested that the APS take the lead in improving these programs by providing the software expertise required to make these codes more accessible to the user. This would involve tasks such as putting a user-friendly interface on analysis programs as well as more complicated projects such as taking an active role in the development and distribution of the most frequently used x-ray analysis packages.

- **Frameworks, Workbenches, and Toolkits**

A common interface program to control all the aspects of the scientific workflow from data acquisition through reduction and analysis was requested by several users. Such an application would enable the semi-automated flow of data from acquisition to reduction and analysis codes. Furthermore, it was perceived that a common interface would enable the novice user to more quickly become proficient and efficient in the use of APS beamlines.

- **Other cited needs**

- Help, documentation, and support
- High-speed data handling and storage
- High-performance computing and remote access
- Common data format

- Collaborative, commonly available, easier-to-use software tools
- Coordination and avoidance of duplication of effort
- A software repository and code maintenance

Reference

A1. P. Jemian, “2006 XSD Scientific Software User Survey,” ANL/APS/TB-52, August 26, 2006.

Appendix 3: Summary of Workshop Presentations

Program

Tuesday, August 29

9:00-9:10	Introduction	Gabrielle Long (APS-XSD)
9:15-9:45	Data Archive and Science Software Tools in Ground-Based and Space Astronomy	Knox Long (Space Telescope Science Institute)
9:50-10:20	A Data Analysis Framework for the Neutron Community	Michael McKerns (Caltech)
10:25-10:45	Visualization and Analysis with ISAW	Tom Worlton (PNS)
10:50-11:10	Break	
11:15-11:35	Software Development at SNS	Steve Miller (SNS)
11:40-12:00	Visualization, Collaboration, and the Grid	Michael E. Papka (MCS)
12:05-12:25	A User's Perspective on Software - Lesson's Learned from DANSE	Ersan Ustundag (Iowa State Univ.)
12:30-1:30	Lunch	
1:35-1:55	NeXus: The Advantages of a Common Data Format	Ray Osborn (MSD)
2:00-2:20	Workbenches, Toolboxes, and Frameworks	Ken Evans (APS-AES)
2:25-2:40	Workflow of the Data Analysis Process	Pete Jemian (APS-AES)
2:45-3:00	Software and Hardware Solution for the Tomography System at Sector 2	Francesco DeCarlo (APS-XSD)
3:05-3:20	Break	
3:25-3:40	Powder Diffraction Crystallography Software: the Present & Future	Brian Toby (APS-XSD)
3:45-4:00	Theory and Modeling	Michel VanVeenendaal (APS-XSD)
4:05-4:20	XAFS Analysis and Modeling	Matt Newville (APS-XSD)
4:25-5:25	Discussion	

Knox Long: Data Archive and Science Software Tools in Ground-Based and Space Astronomy

Knox Long from the Space Telescope Science Institute opened the Workshop with a presentation of space- and ground-based astronomy scientific software tools from the perspective of a user. The motivation for creating standardized data reduction and analysis tools was to address the needs of large numbers of “inexperienced” users on the new space observatories, and to attempt to assure the quality of the emerging science. A file format system called “FITS” (Flexible Image Transport System) was devised to standardize the transport of radio data and CCD images. The basic idea is to have ASCII header records and machine independent binary data storage. The standard data format covers multidimensional arrays (1-D spectra, 2-D images, 3-D+ data cubes), ASCII and binary tables, groups to associate data, and coordinate system mapping (e.g., pixel position and wavelength). This approach has enabled FITS I/O packages for all major languages/operating systems (FORTRAN, C, C++, Python, JAVA, Perl, TCL, Matlab, Mathematica, and IDL) as well as wide interoperability of tools such as image displays and the ability to overlay datasets from many different facilities. Long also described the advantages and disadvantages of early systems (IRAF and a commercial package, IDL), as well modern packages (Pyraf, CIAO, and the Virtual Observatory). His conclusions were that the benefits of standardized data formats across instruments and facilities had the greatest positive impact. If archiving is required, then it is further necessary that data taking modes are standardized, that the metadata be adequate to describe the measurements, and that the analysis becomes standardized. The user base needs to be large and the architecture must provide the functionality to attract them and keep them using and further developing it.

Michael McKerns: A Data Analysis Framework for the Neutron Community

Michael McKerns, Project Manager of the DANSE project at Caltech, presented an overview of DANSE. DANSE is a 12M\$, five-year, NSF IMR-MIP software construction project that is a collaborative effort between software professionals, neutron scattering scientists, and facilities. He described the current situation and presented stereotypes of seven categories of users. He pointed out that Python is a modern object-oriented language that is robust, portable, mature, well-supported, well-documented, easily extendable, and supports rapid application development. It provides scripting and binding to C++, FORTRAN, and others. Through Python, DANSE will have access to many tools. He described the benefits of a framework and discussed in detail the Pyre Framework upon which DANSE is based, including modularity of components, the component data flow paradigm, component implementation, abstraction, the visual interface, distributed/parallel computing, the broad scientific scope, and how it facilitates new and better science. DANSE’s goal is to provide a community-supported open-source software environment for scattering research that, among other things, integrates the capabilities that are available today, provides powerful new applications, provides a coherent framework onto which software components can easily be added by scientists, lowers the barrier for software development, and minimizes duplication of effort. It is

now started with milestones of a beta release in 2008, release 1.0 in 2009, and transition to the community in 2009.

Tom Worlton: Visualization and Analysis with ISAW

Tom Worlton, the Group Leader for Controls and Computing at the Intense Pulsed Neutron Source (IPNS) at Argonne, presented an overview of ISAW. The ISAW project is a collaboration between Argonne and the University of Wisconsin-Stout, started in 1999, that performs data reduction and analysis of neutron scattering data. The ISAW Graphical User Interface (GUI), written in Java, allows the user to read, view, transform, and save data intuitively. It has a scripting capability to handle repetitive processes, with GUI elements that are generated automatically for the specified input parameters. ISAW can also be modified and/or extended by the user through scripts and operators. Worlton presented many slides of ISAW features. He listed lessons learned, including: Java works very well for scientific software; although a GUI is useful for novice users, a scripting language provides an easier way for scientific users to customize an application and automate processes; a combination of a GUI and scripting language is an ideal combination of ease of use and ease of customization; and there is a need for a standardized way to store and retrieve data, such as NeXus/HDF. He pointed out that early ISAW development was restricted to a small team of software developers, but it is now guided through feedback from instrument scientists. He also noted that the ISAW framework easily allows the addition of new operations for data reduction and analysis, with an “Operator Generator” to make this even easier. In addition, new “Wizards” can be constructed from a set of scripts using a “Wizard Generator.” He stated that they would be happy to collaborate on adding software classes to support x-ray data.

Steve Miller: Software Development at SNS

Steve Miller, the Analysis Software Team Leader for the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory, presented the computing and networking environment to be used at SNS. The talk focused on the architecture for storage of, access to, and reduction of the large data sets envisioned for the SNS. Much of the talk focused on the structure of the network and the interactions of various software components for data collection, storage, and retrieval. The SNS system will provide both remote access to the experiment as it happens as well as remote browsing and reduction of data stored on the SNS servers. Miller showed example web portal applications for data retrieval and reduction using a variety of third-party components (DAVE, ISAW, NeXus, and DANSE) for dealing with SNS data. The challenge of ensuring security of data and networks while allowing convenient off-site access to research scientists was also discussed.

Michael E. Papka: Visualization, Collaboration, and the Grid

Michael E. Papka, from the Mathematics and Computer Science Division at Argonne National Laboratory, presented an overview of tools for visualization, collaboration, and grid computing. His slides contained many pictures. For visualization, he indicated that gaming cards bring unprecedented graphics capabilities, and there are high-level

abstractions, such as OpenGL for development. Collaboration is about bringing individuals and groups together in a media-rich environment, and he described the Access Grid Toolkit 3.x as a means to do that. The Grid enables the federation of resources to support applications and communities. It consists of distributed computers, storage, data, and people, with networks to provide connectivity; software and standards to provide the glue; and infrastructure services to facilitate operation. He presented several examples, including the Advanced Biomedical Collaboration (Univ. of Chicago), the FusionGrid (DOE), the Human Arterial Tree Simulation (Brown), and an LDRD project at the APS (ANL, Univ. of Chicago). He ended by listing the many resources available from or through the Mathematics and Computer Science Division.

Ersan Üstündag: A User's Perspective on Software: Lessons Learned from DANSE

Ersan Üstündag, from Iowa State University, presented a user's viewpoint of software for diffraction on engineering materials with particular reference to the neutron diffraction DANSE project. The main objective is a prediction of lifetime and performance of engineering components and materials. Diffraction can provide accurate in-situ constitutive laws and stress measurements under service conditions. Neutron and (high energy) x-ray diffraction are powerful tools of a mostly complementary character. The DANSE project aims to provide a framework enabling interaction of data acquisition, mechanics modeling, and experiment simulation software. Üstündag pointed out that such a framework is crucial to design and direct efficient experiments and therefore exploit the scientific opportunities. He presented several case studies. The main objective of a study of a bulk metallic glass-tungsten composite was the determination of the constitutive laws. In particular power-law and Voce strain hardening were simulated by finite element analysis. Neural network analysis indicated that only a narrow strain region just above the onset of plastic deformation is most significant to distinguish the models. As simulations were not available during the experiment, only few data points were recorded in the important parameter region. He mentioned strain anisotropy analysis and microstructure modeling as particular challenges. He demonstrated dynamic diffraction effects as examples of a specific microstructure. The simulation of diffraction patterns was of central importance in order to understand domain switching in ferroelectric polycrystalline ceramics. Finally, he pointed out that often a combination of techniques, such as diffraction and imaging (tomography), is required and needs to be part of a comprehensive analysis framework.

Ray Osborn: NeXus: The Advantages of a Common Data Format

Ray Osborn is a physicist in the Materials Science Division of Argonne National Laboratory who makes extensive use of both x-ray and neutron scattering techniques for his research. He helped found the NeXus project, which he described in his talk. NeXus is a standardized hierarchical format for recording experimental data and associated metadata. It consists of a growing set of instrument definitions that define how data are recorded for a particular experiment type and a set of software for writing to and retrieving information from NeXus files. The talk presented examples of how NeXus is designed and how it is used. NeXus was designed for both x-ray and neutron needs and

has been endorsed and adopted by nearly all the world's major neutron sources. (For example, the SNS will use NeXus on all its instruments.) There has been interest and involvement from the synchrotron community, but not at the same level as within the neutron community. The talk concluded that use of a common data format allows software to be written for general rather than instrument-specific use and increases productivity of the scientists. NeXus is the only format that is general, has a machine-independent binary encoding, and can handle the wide range of APS instrumentation.

Ken Evans: Workbenches, Toolboxes, and Frameworks

Ken Evans from the APS Controls Group presented examples of common software concepts, such as workbenches (applications from which you can do all or most of your work), toolboxes (groups of applications), and frameworks (collections of routines that you use to create applications) that are candidates for x-ray software. He presented ISAW, GumTree, Java Analysis Studio, and ROOT as examples of workbenches and described the APS Operations and Analysis Group tools as an example of toolboxes. He discussed frameworks, and gave several arguments for using Java, including: it has become a major language owing to its wide use in commercial business applications; its performance approaches that of C; there is already extensive scientific development in Java; and it has features such as “write once, run anywhere” and outstanding development tools. He showed a prototype Java 2-D and 3-D image analysis tool. He pointed out that frameworks written in C /C++ are more amenable to being language agnostic. He described Eclipse, a relatively new, open-source development tool in detail, emphasizing its flexibility, extensibility, and community and industry support. He noted that it can be used for tools as well as a workbench. He indicated the potential for a truly global workbench for both neutron and x-ray data analyses, important because many researchers work at different facilities with both kinds of sources. He concluded that there is a need for all of workbenches, toolboxes, and frameworks; that flexible, extensible, and language agnostic features are important; that open source is desirable; and that collaboration is important.

Pete Jemian: Workflow of the Data Analysis Process

Pete Jemian, the Group Leader for Beamline Controls and Data Acquisition (BCDA) at the APS, presented a discussion of the workflow of data analysis. He described several examples in detail, with diagrams. These included HE-XRD with Mechanical Loading, GSAS, and USAXS data from 33ID-D, as well as an example from the e-Chemistry Laboratory involving grid resources. He presented a diagram of the typical data analysis workflow and expanded on the part involving simulation, noting they have similar processes and requirements. Jemian pointed out that many experiments depart the APS with reduced data, such as I vs. 2θ , whereas others progress to the preliminary model stage during beam time. Further analysis is typically done at the home institution, prolonging the analysis time. A common set of visualization tools, subroutines, utilities (including the software glue code that ties the boxes together), and data format(s) would improve efficiency of the analysis process. Availability of preliminary models during beam time would provide useful feedback to guide experimental choices.

Francesco DeCarlo: Software and Hardware Solution for the Tomography System at Sector 2

Francesco De Carlo, scientist at beamline 2-BM of the APS, gave a presentation on the tomography system developed in sector 2. De Carlo presented X-ray microtomography as an established tool for 3-D imaging of thick structures at micron scale. The fast microtomography system developed at beamline 2-BM offers near video-rate acquisition of tomographic data at micrometer spatial resolution, pipelined processing, and 3-D visualization. At its maximum throughput, the system can image hundreds of specimens per day. Every sample is fully analyzed within 2-3 minutes, giving the user immediate feedback on the quality of the results. The final results are usually presented as a 3-D volume or as slices. The quasi-real-time feature of the system has been instrumental in several applications in both biological applications, where a statistical approach is needed to characterize a broad population of samples, and in material science, where time-dependent 3-D sample evolution can be studied on practical time scales. The entire instrument, including the tomography setup, automatic sample loader, beamline, and a dedicated 32-node computer cluster for data analysis, is also remotely accessible via Access Grid (AG) technology, giving a user full remote control of every aspect of the experiment. Several application examples have been presented including polymer foam modeling, corrosion studies of aluminum, and crack propagation in dental materials. De Carlo pointed out that the data analysis leading to the 3-D rendering is a well-established method. On the other hand, the post-reconstruction data analysis leading to volumetric parameter extraction and/or structure modeling still needs improvements. The tomography program will certainly benefit from an APS-based scientific software development team close to the beamline and user operation.

Brian Toby: Powder Diffraction Crystallography Software: the Present & Future

Brian Toby leads the Materials Characterization group within the APS and is the author of EXPGUI, a graphical user interface to Robert Von Dreele's GSAS crystallographic package. Brian's talk presented how crystallography is used to build atomistic models from diffraction measurements, because we have been able to quantitatively simulate the experiment from the model. We can now do this from many types of measurements, though few offer as many independent observations as does diffraction. For complex contemporary problems, diffraction alone is not sufficient, however, and some contemporary and many future research problems will require the pooling of multiple types of measurements, possibly along with theory. This can not be done with the current generation of crystallographic software. The talk then outlined an initiative that he and Von Dreele have proposed, to create a new generation of crystallographic code that is modular in construction and with an open structure. Such a code would allow scientists to optimize atomistic models using any type of data that can be fitted quantitatively as well as theory or *a priori* information.

Michel van Veenendaal: Theory and Modeling

Michel van Veenendaal, from Northern Illinois University and Argonne, gave a presentation that stressed the importance of theory and modeling. He started by noting that there were quite a few presentations focused on software facilitating science, but few on software actually doing science. One can broadly distinguish two classes of experiments at the APS. First, many experiments mainly focus on structure, such as crystallography (in biology and materials science) or macroscopic studies. The data acquisition and analysis software needs are often substantial here; however, the need for data modeling is less crucial. The obtained structural information is often sufficient for a publication. Another broad class of experiments focuses on excited-state properties, such as electronic, magnetic, and lattice excitations. Often the data acquisition and analysis is straightforward. However, these experiments often require extensive theory and modeling before the data can be published. In addition, even for structural studies, theory and modeling is essential when dealing with, e.g. nanoparticles, fluctuations, dynamics, and non-equilibrium systems. To play a leading role in scientific software development, the APS needs to be present at all levels, including the development of new theory that might eventually lead to new scientific software. Theorists at the APS can play an important role in suggesting and optimizing experiments, interpreting experiments that go beyond conventional models, interfacing with the theoretical community in academia, and forming a sounding board for new scientific ideas. The presence of a theory group is a natural extension of the goal of the APS to maintain its leadership in x-ray science.

Matt Newville: Software for XAFS Modeling and Analysis

Matt Newville, from the University of Chicago, presented the current status and future needs for modeling and analysis of x-ray absorption fine-structure (XAFS). While XAFS shares some algorithms with x-ray scattering, there are unique features of XAFS that make it particularly dependent on high-quality, first-principles calculations of photo-electron scattering in the presence of an excited atom. Currently, these theoretical needs are adequately met by a few codes, notably FEFF, developed by J. Rehr et al. at the University of Washington. Newville showed that the general workflow for XAFS data processing, including visualization, reduction, and quick inspection at the beamline, is adequately handled by GUIs such as Athena from the IFeffit package. This package also includes portable GUIs for full spectral analysis using FEFF and can be used on beamline computers as well as on users' personal machines. While the analysis and visualization tools need some improvements, this task is underway using Python for easier addition of new features. Currently the analysis tools are limited by both the quality and flexibility of the theoretical codes for the photo-electron scattering factors. As new spectroscopies (e.g., RIXS, q-dependent x-ray Raman spectroscopy) are developed and data quality of existing ones (XANES, EXAFS) continues to improve, work is needed on both analysis tools and fundamental theory.